REMARKS/ARGUMENTS

These remarks are made in response to the Office Action of May 15, 2003 (Office Action). As this response is timely filed within the three-month shortened statutory period for reply, no fee is believed due.

In paragraph 2 of the Office Action, claims 1-4, 7-15, and 18-24 have been rejected under 35 U.S.C. § 102(e) as being anticipated by U.S. Patent No. 6,311,159 to Van Tichelen *et al.* (Van Tichelen). In paragraph 4, claims 1-25 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,510,414 to Chaves (Chaves) in view of Van Tichelen.

In response, independent claims 8, 19, and 23 have been amended to clarify that the present invention determines prosodic characteristics of received dual tone multi-frequency (DTMF) signals, groups the DTMF signals, and converts the DTMF signals to textual representations according to the grouping step. Dependent claims 9, 10, 11, 20, 21, and 22 have been amended to clarify that contextual information determined from user utterances can be used to perform the grouping of DTMF signals as well. Claim 26 has been added which reflects the use of contextual information in performing DTMF signal grouping. System claims 27-30 also have been added which reflect the ability of the present invention to convert DTMF signals to text using prosodic information. Please cancel claims 1, 2, 3, 4, 5, 6, 7, 12, 13, 14, 15, 16, 17, and 18 without prejudice. Support for these amendments can be found at page 15, lines 5 – page 16, line 18. No new matter has been added.

Prior to addressing the rejections on the art, a brief review of the Applicant's invention is appropriate. The Applicant has invented a method, system, and apparatus for processing user inputs specifying DTMF signals. In particular, the present invention can analyze prosodic information corresponding to received DTMF signals and group the signals based on the prosodic information. The DTMF signals then can be converted to text equivalents based upon the grouping step.

In illustration, when receiving a series of DTMF signals specifying the digit sequence "102070", the timing between each received digit can be determined. Rather than converting the digit sequence to text as follows: "one", "zero", "two", "zero",

"seven", and "zero", the digits can be grouped according to the time detected between each respective digit. For example, digits received in close succession can be grouped together while detected pauses or longer periods of time can be used to separate digit groupings. In consequence, if pauses are detected as follows: "10" <pause> "20" <pause> "70", the digit sequence can be converted to the text equivalent "ten", "twenty", and "70".

The present invention provides significant advantages when attempting to discern meaning from received user inputs, particularly in the context of natural language understanding. For instance, the present invention can aid in recognizing birthdays and other dates in terms of "month", "date", and "year". The Applicant's invention can use detected pauses to determine that the received string "1" <pause> "26" <pause> "70" corresponds to "one", "twenty-six", "nineteen hundred seventy", rather than "twelve", "six", "seventy" or some other variant which might otherwise result.

The Applicant's invention also incorporates contextual information, such as is determined from a natural language understanding system, to aid in the grouping of DTMF signals. For example, if a date or phone number is expected from the determined context of a user input, this information can be used to group received DTMF digits to represent a date or phone number, rather than determining a string of digits that would otherwise be nonsensical in the determined context.

Turning to the rejections on the art, claims 1-4, 7-15, and 18-24 have been rejected under 35 U.S.C. § 102(e) as being anticipated by Van Tichelen. Van Tichelen discloses a speech controlled computer user interface that can receive speech and DTMF signals. The system converts both speech and DTMF signals to text and derives semantic meaning from the text.

With respect to independent claims 8, 19, and 23, it is asserted that Van Tichelen inherently teaches that prosodic information can be determined from received DTMF signals. In support, column 1, lines 60-67 and column 3, lines 1-7 of the Van Tichelen specification have been cited. Van Tichelen, however, states as follows:

> Column 1, lines 60-67: In a further embodiment, the speech layer may include at least one of: a DTMF module that converts Dial Tone Multi-

> Frequency (DTMF) tones into representative text-based codes; an ASR module that converts speech signals into representative text using Automatic Speech Recognition (ASR) techniques; an SMC module that converts acoustic signals into digitally encoded speech signals using Speech/Music Compression (SMC) techniques; a concatenation module that converts text messages into electronic speech representative signals; and a TTS (Text-to-Speech) module that converts text messages into representative acoustic speech signals.
>
> Column 3, lines 1-7: In a further embodiment, converting between speech messages and text messages may include at least one of: converting Dial Tone Multi-Frequency (DTMF) tones into representative text-based codes with a DTMF module; converting speech signals into representative text using Automatic Speech Recognition (ASR) techniques with an ASR module; converting acoustic signals into digitally encoded speech signals using Speech/Music Compression (SMC) techniques with an SMC module
> . . . .

While Van Tichelen discusses converting both speech and DTMF signals to text in the above passages, Van Tichelen is notably silent with respect to performing any sort of prosodic analysis of received DTMF signals. Moreover, Van Tichelen does not even suggest or imply that such an analysis takes place as noted in the Office Action.

In contrast to the teachings of Van Tichelen, the Applicant's invention analyzes prosodic information relating to received DTMF signals. This information is used to group DTMF signals for purposes of determining textual representations of the grouped DTMF signals. Van Tichelen performs neither a prosodic analysis of DTMF signals, nor a grouping of DTMF signals.

In illustration, upon receiving DTMF signals representing the digit string "1234", Van Tichelen would produce the text "one", "two", "three", and "four". Van Tichelen would be unable to produce variable results of "one" and "two-hundred thirty-four"; "twelve" and "thirty-four"; or "one hundred twenty-three" and "four" based upon the prosodic information determined from the received DTMF signals.

Regarding claims 9, 10, 11, 20, 21, 22, 23, 24, and 25, it is asserted that Van Tichelen inherently teaches that a natural language understanding module provides contextual feedback. In support, column 3, lines 12-26 have been cited. This portion of Van Tichelen states as follows:

> Converting between text messages and semantic meaning messages may include converting, with a natural language understanding module, text messages from the speech layer into representative semantic meaning messages for the discourse layer and/or converting, with a message generator module, semantic meaning messages from the discourse layer into representative text messages for the speech layer.

The above passage illustrates that Van Tichelen teaches a system of determining semantic meaning from text. In particular, text from a speech layer can be provided to a natural language understanding unit, where meaning is determined. The text meaning is provided to a discourse layer. Semantic meaning messages from the discourse layer also can be converted to text messages for the speech layer to be played to the user.

Van Tichelen, however, does not teach that contextual information determined from a natural language understanding system can be used to group DTMF signals and convert those signals to textual representations. In fact, Van Tichelen does not disclose any sort of feedback mechanism between a natural language understanding system and a DTMF converter. Rather, Van Tichelen merely illustrates two-way communication between the user and the dialog system – particularly that meaning can be derived from text and that text can be derived from semantic meaning. In light of the foregoing, withdrawal of the 35 U.S.C. § 102(e) rejection with respect to claims 8-11 and 19-25 is respectfully requested.

Claims 1-25 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Chaves in view of Van Tichelen. Chaves teaches a speech recognition assisted data entry system where caller speech is converted to text. Recognition rules, which correspond to entry fields of a data entry application, can be used to process the text.

With respect to independent claims 8, 19, and 23, it is asserted that Chaves inherently teaches that one or more prosodic characteristics of DTMF signals can be determined. In support, column 2, lines 32-40 and column 4, line 66 – column 5, line 4 have been cited. The cited portions of Chaves have been reproduced below.

> Column 2, lines 32-40: For example, according to one aspect of the present invention, a call center agent may activate a speech recognition

application to receive requested numerical information from a caller. In response to a request for the numerical information from the call center agent, the caller may use a telephone keypad to input the requested numerical information. The numerical information may be received by the speech recognition application and displayed to the call center agent, thereby allowing the call center agent to input the numerical data into a data collection application.

Column 4, line 55 – column 5, line 4: Voice files 40 may comprise recorded phrases or messages used by a call center agent 18 to assist the call center agent 18 in communicating with and acquiring information from a caller 10. For example, voice files 40 may comprise a recorded greeting or a recorded request for information corresponding with a particular data entry field of the data entry application 36. Voice files 40 may be accessed by a call center agent 18, converted to audio signals, and transmitted to a caller 10. For example, voice files 40 may be stored as .wav files or other suitable audio file formats.

Computing system 20 also comprises a speech recognition system 42. Speech recognition system 42 comprises a speech recognition application 44. Speech recognition application 44 comprises systems operable to recognize speech components and dual tone multifrequency (DTMF) signals and convert the speech components and DTMF signals to recognized words or characters, such as text and numerals.

The above passages illustrate that while Chaves receives DTMF signals and converts the signals to text, Chaves makes absolutely no mention of determining prosodic information from DTMF signals. Chaves also is notably silent with respect to performing any sort of grouping of DTMF signals based upon determined prosodic information. As such, Chaves suffers from the same deficiency as Van Tichelen, particularly that neither reference teaches or suggests that prosodic information is determined from DTMF inputs or that such prosodic information can be used to group DTMF signals to aid in converting the signals to text.

The Examiner concedes that Chaves does not teach providing the equivalent text to a natural language understanding system to determine meaning, but asserts that Van Tichelen teaches such a step as well as a natural language understanding module providing contextual feedback. Van Tichelen, however, fails to cure the deficiencies of Chaves. As discussed, Van Tichelen, like Chaves, fails to teach or suggest that

prosodic information is determined from DTMF inputs or that such prosodic information can be used to group DTMF signals to aid in converting the signals to text.

Moreover, contrary to the Examiner's assertion, and as discussed, Van Tichelen does not utilize contextual information in order to group DTMF signals and convert the signals to text. Van Tichelen teaches that semantic meaning can be derived from text and that text can be generated from semantic meaning. In contrast, the present invention describes a process where contextual information, as may be determined by a natural language understanding system, can be fed back through the system for use in grouping and converting DTMF signals to text. In one embodiment of the present invention, the contextual information is provided as feedback to a DTMF converter for just this purpose.

As neither Chaves, Van Tichelen, nor any combination thereof teaches or suggests the features of the present invention as claimed, withdrawal of the 35 U.S.C. § 103(a) rejection regarding claims 8-11 and 19-25 is respectfully requested.

The Applicants believe that this application is now in full condition for allowance, which action is respectfully requested. The Applicants request that the Examiner call the undersigned if clarification is needed on any matter within this Response, or if the Examiner believes a telephone interview would expedite the prosecution of the subject application to completion.

Respectfully submitted,

Date: 8/27/03

Gregory A. Nelson, Registration No. 30,577
Kevin T. Cuenot, Registration No. 46,283
AKERMAN SENTERFITT
222 Lakeview Avenue, Suite 400
Post Office Box 3188
West Palm Beach, FL 33402-3188
Telephone: (561) 653-5000